

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In Re U.S. Patent Application)

Applicant: Tomotaka Matsuda)

Serial No.)

Filed: February 17, 2004)

For: DEVICES FOR INTERPRETING AND)
 RETRIEVING XML DOCUMENTS,)
 METHODS OF INTERPRETING)
 AND RETRIEVING XML)
 DOCUMENTS, AND COMPUTER)
 PRODUCT)

Art Unit:)

I hereby certify that this paper is being deposited with the United States Postal Service as EXPRESS MAIL in an envelope addressed to: Mail Stop PATENT APPLICATION, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on this date.

February 17, 2004
 Date

D. J. Carver
 Express Mail Label No.: EV032731602US

CLAIM FOR PRIORITY

Mail Stop PATENT APPLICATION

Commissioner for Patents

P.O. Box 1450

Alexandria, VA 22313-1450

Dear Sir:

Applicants claim foreign priority benefits under 35 U.S.C. § 119 on the basis of the foreign application identified below:

Japanese Patent Application No. 2003-089126, filed March 27, 2003.

A certified copy of the priority document is enclosed.

Respectfully submitted,

Customer No. 24978

GREER, BURNS & CRAIN, LTD.

February 17, 2004
 300 South Wacker Drive
 Suite 2500
 Chicago, Illinois 60606
 Phone: (312) 360-0080
 Fax: (312) 360-9315

By



Patrick G. Burns
 Registration No. 29,367

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日 2 0 0 3 年 3 月 2 7 日
Date of Application:

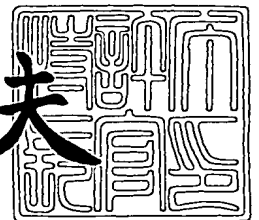
出 願 番 号 特 願 2 0 0 3 - 0 8 9 1 2 6
Application Number:
[ST. 10/C]: [J P 2 0 0 3 - 0 8 9 1 2 6]

出 願 人 富 士 通 株 式 会 社
Applicant(s):

2 0 0 3 年 1 2 月 9 日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



【書類名】 特許願

【整理番号】 0350461

【提出日】 平成15年 3月27日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 17/30

【発明の名称】 XML 文書解析方法、XML 文書検索方法、XML 文書
解析プログラム、XML 文書検索プログラムおよびXML
L 文書検索装置

【請求項の数】 5

【発明者】

 【住所又は居所】 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 富士通
株式会社内

 【氏名】 松田 友隆

【特許出願人】

 【識別番号】 000005223

 【氏名又は名称】 富士通株式会社

【代理人】

 【識別番号】 100104190

 【弁理士】

 【氏名又は名称】 酒井 昭徳

【手数料の表示】

 【予納台帳番号】 041759

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

 【物件名】 図面 1

 【物件名】 要約書 1

 【包括委任状番号】 9906241

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 XML 文書解析方法、XML 文書検索方法、XML 文書解析プログラム、XML 文書検索プログラムおよび XML 文書検索装置

【特許請求の範囲】

【請求項 1】 複数の要素によって所定のツリー構造を構成する XML 文書を入力する XML 文書入力工程と、

前記 XML 文書入力工程によって入力された XML 文書を構成する複数の要素を、要素の出現順に前記ツリー構造にしたがって一列に配列する要素配列工程と、

前記要素配列工程によって一列に配列された各要素内における開始タグと終了タグとの間に記述されているタグ間文字列と、前記開始タグおよび前記終了タグ内に記述されているタグ内文字列と、からなる文字列データを抽出する文字列データ抽出工程と、

前記要素配列工程によって一列に配列された要素ごとに、前記ツリー構造上の種類をあらわすノード型を特定するノード型特定工程と、

前記文字列データ抽出工程によって抽出された文字列データと、前記ノード型特定工程によって特定されたノード型とを前記要素ごとに関連付けるリンクデータを生成するリンクデータ生成工程と、

を含んだことを特徴とする XML 文書解析方法。

【請求項 2】 所定のツリー構造を有する XML 文書に記述されている文字列データの中から、検索対象となる文字列を検索する XML 文書検索方法であって、

前記検索対象となる文字列を検索するための検索条件を入力する検索条件入力工程と、

前記 XML 文書を構成する複数の要素の出現順に前記ツリー構造にしたがって配列された第 1 の文書構造データが、前記検索条件に合致するか否かを判定する第 1 の文書構造データ判定工程と、

前記第 1 の文書構造データ判定工程によって前記検索条件に合致すると判定された場合にのみ、前記検索条件に合致した第 1 の文書構造データに関連付けられ

ている前記文字列データ内の文字列を抽出する第1の関連文字列抽出工程と、

前記第1の関連文字列抽出工程によって抽出された文字列が、前記検索条件に合致するか否かを判定する関連文字列判定工程と、

前記関連文字列判定工程によって前記検索条件に合致すると判定された場合にのみ、前記第1の文書構造データの次に配列されている第2の文書構造データが、前記検索条件に合致するか否かを判定する第2の文書構造データ判定工程と、

前記第2の文書構造データ判定工程によって前記検索条件に合致すると判定された場合にのみ、前記検索条件に合致した第2の文書構造データに関連付けられている前記文字列データ内の文字列を抽出する第2の関連文字列抽出工程と、

を含んだことを特徴とするXML文書検索方法。

【請求項3】 複数の要素によって所定のツリー構造を構成するXML文書を入力させるXML文書入力工程と、

前記XML文書入力工程によって入力されたXML文書を構成する複数の要素を、要素の出現順に前記ツリー構造にしたがって一列に配列させる要素配列工程と、

前記要素配列工程によって一列に配列された各要素内における開始タグと終了タグとの間に記述されているタグ間文字列と、前記開始タグおよび前記終了タグ内に記述されているタグ内文字列と、からなる文字列データを抽出させる文字列データ抽出工程と、

前記要素配列工程によって一列に配列された要素ごとに、前記ツリー構造上の種類をあらわすノード型を特定させるノード型特定工程と、

前記文字列データ抽出工程によって抽出された文字列データと、前記ノード型特定工程によって特定されたノード型とを前記要素ごとに関連付けるリンクデータを生成させるリンクデータ生成工程と、

をコンピュータに実行させることを特徴とするXML文書解析プログラム。

【請求項4】 所定のツリー構造を有するXML文書に記述されている文字列データの中から、検索対象となる文字列を検索させるXML文書検索プログラムであって、

前記検索対象となる文字列を検索するための検索条件を入力させる検索条件入

力工程と、

前記XML文書を構成する複数の要素の出現順に前記ツリー構造にしたがって配列された第1の文書構造データが、前記検索条件に合致するか否かを判定させる第1の文書構造データ判定工程と、

前記第1の文書構造データ判定工程によって前記検索条件に合致すると判定された場合にのみ、前記検索条件に合致した第1の文書構造データに関連付けられている前記文字列データ内の文字列を抽出させる第1の関連文字列抽出工程と、

前記第1の関連文字列抽出工程によって抽出された文字列が、前記検索条件に合致するか否かを判定させる関連文字列判定工程と、

前記関連文字列判定工程によって前記検索条件に合致すると判定された場合にのみ、前記第1の文書構造データの次に配列されている第2の文書構造データが、前記検索条件に合致するか否かを判定させる第2の文書構造データ判定工程と、

前記第2の文書構造データ判定工程によって前記検索条件に合致すると判定された場合にのみ、前記検索条件に合致した第2の文書構造データに関連付けられている前記文字列データ内の文字列を抽出させる第2の関連文字列抽出工程と、

をコンピュータに実行させることを特徴とするXML文書検索プログラム。

【請求項5】 所定のツリー構造を有するXML文書に記述されている文字列データの中から、検索対象となる文字列を検索するXML文書検索装置であって、

前記検索対象となる文字列を検索するための検索条件を入力する検索条件入力手段と、

前記XML文書を構成する複数の要素の出現順に前記ツリー構造にしたがって配列された複数の文書構造データが、前記検索条件にそれぞれ合致するか否かを判定する文書構造データ判定手段と、

前記文書構造データ判定手段によって前記検索条件に合致すると判定された場合にのみ、前記検索条件に合致した文書構造データに関連付けられている前記文字列データ内の文字列を抽出する関連文字列抽出手段と、

前記関連文字列抽出手段によって抽出された文字列が、前記検索条件に合致す

るか否かを判定する関連文字列判定手段と、を備え、

前記関連文字列抽出手段は、前記関連文字列判定手段によって抽出された文字列が、前記検索条件に合致すると判定され、かつ、前記文書構造データ判定手段によって、前記検索条件に合致すると判定された文書構造データの次に配列されている文書構造データが前記検索条件に合致すると判定された場合、前記検索条件に合致すると判定された文書構造データ以外の文書構造データに対して関連付けられている前記文字列データ内の文字列を、前記検索対象となる文字列として抽出することを特徴とするXML文書検索装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

この発明は、XML文書の解析や解析済みのXML文書の検索をおこなうXML文書解析方法、XML文書検索方法、XML文書解析プログラム、XML文書検索プログラムおよびXML文書検索装置に関する。

【0002】

【従来の技術】

XMLは、周知の通り、タグによって情報を埋め込むことができるマークアップ言語であり、このタグによって作成されたXML文書は、文書構造と書式情報が厳格に分離されているため、HTMLにかわってマークアップ言語の主流となっている。

【0003】

このように、シンプルな文書構造を有するXML文書の解析には、通常、XMLパーサが用いられている。XMLパーサは、XML文書を読み込み、読み込んだXML文書がどのような文書構造となっているかを解析する機能を有するモジュールである。

【0004】

このXMLパーサが採用するAPI (Application Programming Interface) には、DOM (Document Object Model) と、SAX (Simple API for XML) と、

が標準化されており、これらの A P I を実装した X M L パーサの利用が主流になっている。

【 0 0 0 5 】

一般的な D O M パーサは、最初に X M L 文書全体を読み込み、X M L 文書を構成している要素、属性その他の文字データを、ツリー構造として解析するよう実装されている。また、一般的な S A X パーサは、D O M とは異なり、X M L 文書全体を読み込まず、X M L 文書の先頭、すなわち、要素の出現順から順に解析するよう実装されている。

【 0 0 0 6 】

また、X M L 文書を検索するものとしては、意味的に類似した D T D をもつ X M L データベースに対して、ユーザが D T D の差異を意識せずに効率的に検索することのできる X M L 文書検索装置が提案されている（たとえば、下記特許文献 1 参照。）。

【 0 0 0 7 】

【特許文献 1】

特開 2 0 0 0 - 2 5 0 9 3 8 号公報

【 0 0 0 8 】

【発明が解決しようとする課題】

しかしながら、あらかじめ X M L 文書が格納されているデータベースから、X M L 文書の抽出または検索を行う場合は、データベース内の X M L 文書を解析してから、該当する X M L 文書を抽出または検索する必要がある。ここで、上述のような一般的な D O M パーサによる解析処理では、X M L 文書全体を読み込まなければならないため、データベース内の X M L 文書すべての解析処理をおこなうと、X M L 文書を抽出または検索するまで多大な時間がかかることとなる。

【 0 0 0 9 】

また、上述の一般的な S A X パーサでは、X M L 文書全体を読み込む必要はないが、タグ内に埋め込まれた情報やタグ間に挟まれたすべての情報を読み込む必要があるため、データベース内の X M L 文書すべての解析処理をおこなうと、X M L 文書を抽出または検索するまで多大な時間がかかることとなる。

【0010】

また、特許文献1のXML文書検索装置では、データベースクライアントによって作成された入力用検索式により、入力解析部でその要素名を抽出し、抽出された要素名の類義語を類義語抽出部から取得し、類義語とカテゴリ類推部に格納されている要素名とを比較し、一致する要素名を選択しているため、検索に際し、すべての要素名を比較対象としていることから、検索に時間がかかることとなる。

【0011】

この発明は、上述した従来技術による問題点を解消するため、XML文書の検索の高速化を図ることができるXML文書解析方法、XML文書検索方法、XML文書解析プログラム、XML文書検索プログラムおよびXML文書検索装置を提供することを目的とする。

【0012】**【課題を解決するための手段】**

上述した課題を解決し、目的を達成するため、この発明にかかるXML文書解析方法、XML文書解析プログラムおよびXML文書解析装置は、複数の要素によって所定のツリー構造を構成するXML文書を入力し、前記XML文書入力工程によって入力されたXML文書を構成する複数の要素を、要素の出現順に前記ツリー構造にしたがって一列に配列し、この配列によって一列に配列された各要素内における開始タグと終了タグとの間に記述されているタグ間文字列と、前記開始タグおよび前記終了タグ内に記述されているタグ内文字列と、からなる文字列データを抽出し、一列に配列された要素ごとに、前記ツリー構造上の種類をあらわすノード型を特定し、前記文字列データ抽出工程によって抽出された文字列データと特定されたノード型とを前記要素ごとに関連付けるリンクデータを生成することを特徴とする。

【0013】

また、上記発明は、分岐されている子要素を、その分岐元の要素によって挟み込むことにより、前記ツリー構造にしたがって一列に配列することとしてもよい。この発明によれば、XML文書を高速に検索することができる解析済XML文

書を生成することができる。

【0014】

また、この発明にかかるXML文書検索方法、XML文書検索プログラムおよびXML文書検索装置は、所定のツリー構造を有するXML文書に記述されている文字列データの中から、検索対象となる文字列を検索するための検索条件を入力し、前記XML文書を構成する複数の要素の出現順に前記ツリー構造にしたがって配列された第1の文書構造データが、前記検索条件に合致するか否かを判定し、この判定によって前記検索条件に合致すると判定された場合にのみ、前記検索条件に合致した第1の文書構造データに関連付けられている前記文字列データ内の文字列を抽出し、この抽出によって抽出された文字列が、前記検索条件に合致するか否かを判定し、この判定によって前記検索条件に合致すると判定された場合にのみ、前記第1の文書構造データの次に配列されている第2の文書構造データが、前記検索条件に合致するか否かを判定し、この判定によって前記検索条件に合致すると判定された場合にのみ、前記検索条件に合致した第2の文書構造データに関連付けられている前記文字列データ内の文字列を抽出することを特徴とする。

【0015】

また、上記発明において、前記第1および第2の文書構造データは、前記ツリー構造を構成するノードの種類をあらわすノード型と、前記文字列データに対し関連付けするリンクデータとを含むこととしてもよい。さらに、文書構造データを判定する場合、前記ノード型および前記リンクデータが検索条件に合致するか否かを判定することとしてもよい。また、文書構造データと関連付けられている文字列を抽出する場合は、前記リンクデータに基づいて、前記文字列データを抽出することとしてもよい。

【0016】

この発明によれば、検索条件と文書構造データとの判定をおこない、検索条件と文書構造データとが合致する場合にのみ文字列データを抽出する。換言すれば、文書構造データが検索条件に合致しない場合は、文字列データを参照することができない。これにより、文字列データの参照頻度を少なくすることができる。

【0017】**【発明の実施の形態】**

以下に添付図面を参照して、この発明にかかるXML文書解析方法、XML文書検索方法、XML文書解析プログラム、XML文書検索プログラムおよびXML文書検索装置の好適な実施の形態を詳細に説明する。

【0018】

(XML文書解析／検索装置のハードウェア構成)

まず、図1を用いて、本実施の形態にかかるXML文書解析／検索装置のハードウェア構成について説明する。XML文書解析／検索装置は、CPU101と、ROM102と、RAM103と、HDD(ハードディスクドライブ)104と、HD(ハードディスク)105と、FDD(フレキシブルディスクドライブ)106と、着脱可能な記録媒体の一例としてFD(フレキシブルディスク)107と、ディスプレイ108と、I/F(インターフェース)109と、キーボード110と、マウス111と、スキャナ112と、プリンタ113と、を備えている。また、各構成部はバス100によってそれぞれ接続されている。

【0019】

ここで、CPU101は、XML文書解析／検索装置の全体の制御を司る。ROM102は、ブートプログラムなどのプログラムを記憶している。RAM103は、CPU101のワークエリアとして使用される。HDD104は、CPU101の制御にしたがってHD105に対するデータのリード／ライトを制御する。HD105は、HDD104の制御で書き込まれたデータを記憶する。

【0020】

FDD106は、CPU101の制御にしたがってFD107に対するデータのリード／ライトを制御する。FD107は、FDD106の制御で書き込まれたデータを記憶したり、FD107に記憶されたデータをXML文書解析／検索装置に読み取らせたりする。着脱可能な記録媒体として、FD107のほか、CD-ROM(CD-R、CD-RW)、MO、DVD(Digital Versatile Disk)、メモリーカードなどであってもよい。ディスプレイ108は、カーソル、アイコンあるいはツールボックスをはじめ、文書、画像、機能情報などのデータを

表示する。このディスプレイ 108 は、たとえば、CRT、TFT 液晶ディスプレイ、プラズマディスプレイなどを採用することができる。

【0021】

I/F 109 は、通信回線を通じてインターネットなどのネットワークに接続され、このネットワークを介して他の装置に接続される。そして、I/F 109 は、ネットワークと内部のインターフェースを司り、外部装置からのデータの入出力を制御する。I/F 109 には、たとえばモデムや LAN アダプタなどを採用することができる。

【0022】

キーボード 110 は、文字、数字、各種指示などの入力のためのキーを備え、データの入力をおこなう。また、タッチパネル式の入力パッドやテンキーなどであってもよい。マウス 111 は、カーソルの移動や範囲選択、あるいはウィンドウの移動やサイズの変更などをおこなう。ポインティングデバイスとして同様に機能を備えるものであれば、トラックボールやジョイスティックなどであってもよい。

【0023】

スキャナ 112 は、画像を光学的に読み取り、XML 文書解析／検索装置内に画像データを取り込む。また、プリンタ 113 は、画像データや文書データを印刷する。プリンタ 113 には、たとえば、レーザプリンタやインクジェットプリンタを採用することができる。

【0024】

(XML 文書解析／文書検索装置の機能的構成)

つぎに、XML 文書解析／検索装置 200 の機能的構成について、図 2 を用いて説明する。図 2 は、XML 文書解析／検索装置 200 の機能的構成を示すブロック図である。

【0025】

XML 文書解析／検索装置 200 は、XML 文書入力部 201 と、XML 文書解析部 202 と、XML 文書格納部 208 と、検索条件入力部 211 と、検索データ作成部 212 と、XML 文書検索部 213 と、検索結果出力部 218 と、か

ら構成されている。

【0026】

XML文書入力部201は、図1に示すキーボード110やマウス111などの入力装置から、保存すべきXML文書を入力する。この入力すべきXML文書は、図3に示すように、先頭の要素から分岐したツリー構造とされている。ここで要素1は、ルート要素であり、要素2～4は、ルート要素1の子要素である。また、要素5～7は、要素4の子要素である。なお、図3の要素に付されている番号は、XML文書における要素の出現順を示す。このXML文書入力部201は、具体的には、たとえば図1に示したI/F109によってその機能を実現する。

【0027】

XML文書解析部202は、要素配列部203と、文字列データ抽出部204と、ノード型特定部205と、リンクデータ生成部206と、から構成されている。要素配列部203は、XML文書入力部201によって入力されたXML文書を構成する要素の出現順に、そのツリー構造にしたがって複数の要素を一行に配列する。

【0028】

文字列データ抽出部204は、要素配列部203によって一行に配列された各要素から、各要素内における開始タグと終了タグとの間に記述されているタグ間文字列と、開始タグおよび終了タグ内に記述されているタグ内文字列と、からなる文字列データを抽出する。

【0029】

ここで、文字列データは、タグ内に埋め込まれるタグ内文字列およびタグに挟まれたタグ間文字列とからなるデータであり、タグ内文字列とは、たとえば、タグ内に埋め込まれる要素名、属性、名前空間名その他の文字データなどである。また、タグ間文字列とは、開始タグと終了タグとの間に埋め込まれた要素内容その他の文字データなどである。

【0030】

この文字列データは、タグ内文字列やタグ間文字列の長さ（ビット長）が可変

長なデータである。また、文字列データは、名前空間データと、要素名データと、要素内容または属性をあらわす要素内容／属性データと、から構成されている。また、要素内容／属性データは、そのデータの種類によって分けられる。たとえば、そのデータの種類が、テキスト型であるテキストデータと、整数型である整数データと、に分けられる。

【0031】

ノード型特定部205は、要素配列部203によって一列に配列された要素ごとに、そのツリー構造上の種類をあらわすノード型を特定する。ここで、ノード型とは、各要素の構造上の種類をあらわすデータであり、たとえば、XML文書の先頭位置であることを示す「ドキュメント開始」、XML文書の終了位置であることを示す「ドキュメント終了」、分岐を有する要素（親要素）のうち開始要素であることをあらわす「要素開始」、分岐を有する要素（親要素）のうち閉じ要素であることをあらわす「要素終了」、その要素に名前空間が埋め込まれていることをあらわす「名前空間マップ」、その要素に属性が埋め込まれていることをあらわす「属性」、その要素が分岐した要素であることをあらわす「1子要素」がある。

【0032】

リンクデータ生成部206は、要素ごとに、文字列データ抽出部204によって抽出された文字列データと、ノード型特定部205によって特定されたノード型と、を関連付けるリンクデータを生成する。

【0033】

ここで、リンクデータとは、ノード型と文字列データとを関連付けるIDである。リンクデータは、たとえば、上述の文字列データの中の名前空間データとリンクする名前空間IDと、要素名データとリンクする要素名IDと、要素内容／属性データの種類とリンクするデータ型IDと、要素名／属性データとリンクする要素名／属性データIDと、から構成される。

【0034】

なお、このXML文書解析部202は、具体的には、たとえば図1に示したROM102、RAM103、HD105、FD107などに格納されたプログラ

ムをCPU101が実行することによって、また図1に示したI/F109によってその機能を実現する。

【0035】

XML文書格納部208は、XML文書解析部202によって解析された文書構造データと文字列データとを格納する。このXML文書格納部208は、具体的には、たとえば図1に示したRAM103、HD105、FD107などによってその機能を実現する。

【0036】

また、検索条件入力部211は、図1に示すキーボード110やマウス111などの入力装置やネットワークから入力されてくる検索条件を入力する。この検索条件入力部211は、具体的には、たとえば図1に示したROM102、RAM103、HD105、FD107などに格納されたプログラムをCPU101が実行することによって、また図1に示したI/F109によってその機能を実現する。

【0037】

検索データ作成部212は、たとえばW3Cで勧告されているXPath、XQLなどに基づいて、入力された検索条件から検索データを作成する。この検索データ作成部212は、具体的には、たとえば図1に示したROM102、RAM103、HD105、FD107などに格納されたプログラムをCPU101が実行することによって、また図1に示したI/F109によってその機能を実現する。XML文書検索部213は、検索データに基づいて、XML文書格納部208に格納されているXML文書から、検索データに合致するXML文書内の文字列を検索する。

【0038】

このXML文書検索部213は、ノード型走査部214と、文書構造データ判定部215と、関連文字列抽出部216と、関連文字列判定部217と、から構成されている。ノード型走査部214は、XML文書格納部208に格納された文書構造データのノード型を走査する。

【0039】

文書構造データ判定部 215 は、ノード型走査部 214 によって走査される都度、文書構造データが、検索データに合致するか否かを判定する。また、関連文字列判定部 217 によって検索データに合致すると判定された場合にのみ、走査された文書構造データの次に配列されている文書構造データが、検索データに合致するか否かを判定する。

【0040】

関連文字列抽出部 216 は、文書構造データ判定部 215 によって検索データに合致すると判定された場合にのみ、検索データに合致した文書構造データに関連付けられている文字列データ内の文字列を抽出する。関連文字列判定部 217 は、関連文字列抽出部 216 によって抽出された文字列が、検索データに合致するか否かを判定する。

【0041】

この XML 文書検索部 213 は、具体的には、たとえば図 1 に示した ROM 102、RAM 103、HD 105、FD 107 などに格納されたプログラムを CPU 101 が実行することによって、また図 1 に示した I/F 109 によってその機能を実現する。

【0042】

検索結果出力部 218 は、XML 文書によって検索された文字列（検索対象文字列）を検索結果として出力する。この検索結果出力部 218 は、具体的には、たとえば図 1 に示した I/F 109 によってその機能を実現する。

【0043】

（XML 文書解析方法）

つぎに、XML 文書解析方法について、図 4 のフローチャートを用いて説明する。まず、XML 文書入力部 201 に図 3 に示すツリー構造からなる XML 文書が入力された場合（ステップ S401：Yes）、図 5 に示すように、XML 文書を、先頭から出現する要素（ノード単位）順に配列する（ステップ S402）。具体的には、図 5 に示すように、分岐された子要素の親となる要素は、その分岐された子要素を挟み込む。

【0044】

たとえば、ルート要素 1 から分岐されている子要素（要素 2 ～要素 8）は、ルート要素 1（開始要素 1 A）と閉じ要素 1 Bとによって挟み込まれる。また、要素 4 から分岐されている子要素（要素 5 ～要素 7）は、要素 4（開始要素 4 A）と閉じ要素 4 Bとによって挟み込まれる。これにより、XML 文書を要素の出現順に一系列に配列しても、XML 文書のツリー構造を維持することができる。

【0 0 4 5】

つぎに、一系列に配列された各要素から、各要素内における開始タグと終了タグとの間に記述されているタグ間文字列と、開始タグおよび終了タグ内に記述されているタグ内文字列と、からなる文字列データを抽出する（ステップ S 4 0 3）。

【0 0 4 6】

つぎに、一系列に配列された XML 文書の各要素のノード型を特定する（ステップ S 4 0 4）。また、ノード型と文字列データとを関連付けるリンクデータを生成する（ステップ S 4 0 5）。

【0 0 4 7】

また、一系列に配列された XML 文書は、図 6 に示すように文字列データが抽出されると、ノード型とリンクデータからなる文書構造データとなる。そして、文書構造データと文字列データとを、解析済 XML 文書として XML 文書格納部 2 0 8 に格納する（ステップ S 4 0 6）。

【0 0 4 8】

なお、ここで、図 7 に示す XML 文書を解析して格納した解析済 XML 文書の格納例を、図 8 ～図 1 2 に示す。図 8 は XML 文書格納部 2 0 8 に格納された文書構造データを示している。図 8 に示す各文書構造データのビット長は、たとえば 3 2 ビットまたは 6 4 ビットなどの固定長である。この固定ビット幅を 6 4 ビットとした場合は、ノード型、名前空間 ID、要素名 ID およびデータ型 ID で 3 2 ビットを占める。残りの 3 2 ビットは、要素名／属性データ ID となる。また、図 9 ～図 1 2 は XML 文書格納部 2 0 8 に格納された文字列データを示している。

【0 0 4 9】

(XML文書の検索方法)

つぎに、図8～図12に示したXML文書の検索方法について説明する。図13はXML文書の検索方法の処理手順を示すフローチャートである。まず、検索条件が入力された場合（ステップS1301：Yes）、検索データを作成する（ステップS1302）。図14に検索条件の一例を示す。この検索条件は、名前空間0に属する製品の型番が“0002”である名前（製品名）を検索するための検索条件である。検索データは、入力された検索条件から作成される。図15に、作成された検索データの一例を示す。

【0050】

この検索データは、第1検索データPと、第2検索データQと、第3検索データRと、第4検索データSと、第5検索データTと、から構成されている。第1検索データPおよび第2検索データQは、それぞれ、ノード型と、名前空間IDと、要素名IDと、を検索するノードテスト用データである。

【0051】

また、第3検索データRは、演算子、たとえば、「=」、「≠」、「<」、「>」、「≤」、「≥」をあらわすデータである。第4検索データSは、評価対象をあらわすデータである。第5検索データTは、検索データP、Q、R、Sによる検索をおこなったときのリターンバリューを受けることにより検索されるデータである。

【0052】

この検索データによれば、第1検索データPおよび第2検索データQに該当する評価値（文字列データ）と、第4検索データSの評価対象（評価値と同種の文字列データ）が、第3検索データRの演算子を満足し、かつ第5検索データTが一致する場合は、検索対象となる文字列がXML文書内に存在することとなる。

【0053】

つぎに、作成された検索データにより、XML文書の検索を実行する（ステップS1303）。そして、検索結果リストを出力する（ステップS1304）。ここで、図16を用いて検索実行処理について説明する。まず、検索結果リストを初期化し（ステップS1601）、検索状態を初期化する（ステップS160

2)。つぎに、ノード型を先頭から走査する（ステップS1603）。

【0054】

そして、走査されたノード型について、ノードテストを実行する（ステップS1604）。ノード型が一致しない場合は（ステップS1604：No）、つぎのノード型を走査する（ステップS1603）。ノード型が一致する場合は（ステップS1604：Yes）、名前空間IDが一致するか否かを判定する（ステップS1605）。

【0055】

名前空間IDが一致しない場合は（ステップS1605：No）、つぎのノード型を走査する（ステップS1603）。名前空間IDが一致する場合は（ステップS1605：Yes）、要素名IDが一致するか否かを判定する（ステップS1606）。要素名IDが一致しない場合は（ステップS1606：No）、つぎのノード型を走査する（ステップS1603）。要素名IDが一致する場合は（ステップS1606：Yes）、ノードテストが終了したか否かを判定する（ステップS1607）。

【0056】

ノードテストが終了していない場合は（ステップS1607：No）、ステップS1603に移行する。ノードテストが終了した場合には（ステップS1607：Yes）、評価対象の判定を実行する（ステップS1608）。この後、リターンバリューのノード型が一致するか否かを判定する（ステップS1609）。リターンバリューのノード型が一致しない場合は（ステップS1609：No）、つぎのノード型を走査する（ステップS1603）。

【0057】

このあと、ノード型が一致する場合は（ステップS1609：Yes）、名前空間IDが一致するか否かを判定する（ステップS1610）。名前空間IDが一致しない場合は（ステップS1610：No）、つぎのノード型を走査する（ステップS1603）。名前空間IDが一致する場合は（ステップS1610：Yes）、要素名IDが一致するか否かを判定する（ステップS1611）。

【0058】

要素名IDが一致しない場合は（ステップS1611：No）、つぎのノード型を走査する（ステップS1603）。要素名IDが一致する場合は（ステップS1611：Yes）、評価結果が真でなければ（ステップS1612：No）、状態初期化する（ステップS1602）。評価結果が真であれば（ステップS1612：Yes）、文字列データからステップS1611で一致した要素名IDに対応する要素名を抽出する（ステップS1613）。

【0059】

そして、抽出した要素名を検索結果として検索結果リストに追加する（ステップS1614）。つぎのノード型が「ドキュメント終了」でない場合は（ステップS1615：No）、ステップS1602に移行する。一方、つぎのノード型が「ドキュメント終了」である場合は（ステップS1615：Yes）、検索を終了する。

【0060】

この検索実行処理によれば、ステップS1603～ステップS1607およびステップS1609～ステップS1611、S1615については文書構造データのための検索であり、ステップS1608、ステップS1612およびステップS1613のみが文字列データを含む文字列データの検索であるため、文字列データの参照頻度を少なくすることができる。したがって、検索したい文字列を高速に抽出することができる。

【0061】

つぎに、図8～図12に示すXML文書を、図15の検索データによって検索実行したときの状態遷移について、図17を用いて説明する。なお、図17中の①～⑤は、図15の検索データP、Q、R、S、Tの検索①～⑤に対応する。まず、検索結果リストおよび状態の初期化をおこなう。つぎに、先頭のノード型から第1検索データとノードテストを実行する。第1検索データのノード型は、「要素開始」であるため、文書構造データのノード型が「要素開始」となるまで、ノード型を走査する。

【0062】

そして、ノード型：「要素開始」が走査された場合（走査A）、リンクデータ

の比較をおこなう。この要素名IDは、第1検索データの要素名IDと一致しないため、ノード型の走査を継続する。つぎのノード型ID：「要素開始」が走査された場合、リンクデータの比較をおこなう。この「要素開始」のリンクデータは、第1検索データと一致し、ノードテスト真となる。

【0063】

第1検索データのノードテストが終了したため、つぎのノード型から、第2検索データを用いてノードテストをおこなう。第2検索データのノード型は「属性」であるため、文書構造データのノード型が「属性」となるまで、ノード型を走査する。

【0064】

つぎにノード型：「属性」が走査されると（走査B）、リンクデータの比較をおこなう。この「属性」のリンクデータは、第2検索データと一致し、ノードテスト真となる。ここで、ノードテストが終了となり、第3検索データ（演算子）と第4検索データ（評価対象）の評価をおこなう。

【0065】

すなわち、ノードテストが終了した時点で走査されていたノード型：「属性」のデータ型IDと要素名／属性データIDを参照して、文字列データから対応する文字列データを抽出する。この場合、データ型IDは「テキスト」であり、「要素名／属性データID」は「1」であるため、図12に示す文字列データから文字列データ（テキストデータ）「0001」を抽出する。

【0066】

この抽出した文字列データ「0001」と評価対象「0002」では、第3検索データの演算子「=」を満足しないため、評価結果は偽となる。ノードテストが終了しているため、つぎのノード型：「1子要素」について第5検索データであるリターンバリューのノードテストを実行する（走査C）。

【0067】

このノードテストは真となり、一回目の検索データによる検索が終了する。そして、検索結果を判定する。今回の検索結果は、評価対象の評価が偽であるため、検索不成立となる（走査D）。そして、状態初期化をおこない、つぎのノード

型を走査する。

【0068】

そして、ノード型：「要素開始」が走査された場合（走査E）、リンクデータの比較をおこなう。この「要素開始」のリンクデータは、第1検索データと一致し、ノードテスト真となる。

【0069】

第1検索データのノードテストが終了したため、つぎのノード型から、第2検索データを用いてノードテストをおこなう。第2検索データのノード型は「属性」であるため、文書構造データのノード型が「属性」となるまで、ノード型を走査する。

【0070】

つぎにノード型：「属性」を走査すると（走査F）、リンクデータの比較をおこなう。この「属性」のリンクデータは、第2検索データと一致し、ノードテスト真となる。ここで、ノードテストが終了となり、第3検索データ（演算子）と第4検索データ（評価対象）の評価をおこなう。

【0071】

すなわち、ノードテストが終了した時点で走査されていたノード型：「属性」のデータ型IDと要素名／属性データIDを参照して、文字列データから対応する文字列データを抽出する。この場合、データ型IDは「テキスト」であり、「要素名／属性データID」は「3」であるため、図12に示す文字列データから文字列データ（テキストデータ）「0002」を抽出する。

【0072】

この抽出した文字列データ「0002」と評価対象「0002」では、第3検索データの演算子「=」を満足するため、評価結果は真となる。ノードテストが終了しているため、つぎのノード型：「1子要素」について第5検索データであるリターンバリューのノードテストを実行する（走査G）。

【0073】

このノードテストは真となり、二回目の検索データによる検索が終了する。そして、検索結果を判定する。今回の検索結果は、評価対象の評価が真であり、か

つりターンバリューのノードテストも真であるため、検索成立となる（走査H）。これにより、文字列データから、リターンバリューが真とされたノード型のデータ型ID「テキスト」と要素名／属性データID「0002」に対応する文字列データ「LOOX」を抽出する（図12参照）。

【0074】

そして、この抽出した文字列データ「LOOX」を検索結果リストに追加する。このあと、再度状態初期化をおこない、ノード型の走査を続ける。そして、走査したノード型が「ドキュメント終了」となった場合、終了する。

【0075】

以上説明したように、本実施の形態にかかるXML文書解析／検索装置200は、予めXML文書を解析し、固定長データからなる文書構造データと文字列データからなる文字列データとに分けて格納しておくことができる。また、本実施の形態にかかるXML文書解析／検索装置200は、ビット幅の短い固定長な文書構造データの参照頻度を多くして、可変長な文字列データの参照をノードテストが終了した場合および評価結果をおこなう場合にのみおこなっているため、可変長な文字列データの参照頻度を少なくすることができる。これにより、XML文書内の情報を高速に検索することができる。

【0076】

なお、本実施の形態で説明したXML文書解析方法およびXML文書検索方法は、予め用意されたプログラムをパーソナル・コンピュータやワークステーション等のコンピュータで実行することにより実現することができる。このプログラムは、ハードディスク、フレキシブルディスク、CD-ROM、MO、DVD等のコンピュータで読み取り可能な記録媒体に記録され、コンピュータによって記録媒体から読み出されることによって実行される。またこのプログラムは、インターネット等のネットワークを介して配布することが可能な伝送媒体であってもよい。

【0077】

（付記1）複数の要素によって所定のツリー構造を構成するXML文書を入力するXML文書入力工程と、

前記XML文書入力工程によって入力されたXML文書を構成する複数の要素を、要素の出現順に前記ツリー構造にしたがって一列に配列する要素配列工程と、

前記要素配列工程によって一列に配列された各要素内における開始タグと終了タグとの間に記述されているタグ間文字列と、前記開始タグおよび前記終了タグ内に記述されているタグ内文字列と、からなる文字列データを抽出する文字列データ抽出工程と、

前記要素配列工程によって一列に配列された要素ごとに、前記ツリー構造上の種類をあらわすノード型を特定するノード型特定工程と、

前記文字列データ抽出工程によって抽出された文字列データと、前記ノード型特定工程によって特定されたノード型とを前記要素ごとに関連付けるリンクデータを生成するリンクデータ生成工程と、

を含んだことを特徴とするXML文書解析方法。

【0078】

(付記2) 前記要素配列工程は、分岐されている子要素を、その分岐元の要素によって挟み込むことにより、前記ツリー構造にしたがって一列に配列することを特徴とする付記1に記載のXML文書解析方法。

【0079】

(付記3) 所定のツリー構造を有するXML文書に記述されている文字列データの中から、検索対象となる文字列を検索するXML文書検索方法であって、

前記検索対象となる文字列を検索するための検索条件を入力する検索条件入力工程と、

前記XML文書を構成する複数の要素の出現順に前記ツリー構造にしたがって配列された第1の文書構造データが、前記検索条件に合致するか否かを判定する第1の文書構造データ判定工程と、

前記第1の文書構造データ判定工程によって前記検索条件に合致すると判定された場合にのみ、前記検索条件に合致した第1の文書構造データに関連付けられている前記文字列データ内の文字列を抽出する第1の関連文字列抽出工程と、

前記第1の関連文字列抽出工程によって抽出された文字列が、前記検索条件に

合致するか否かを判定する関連文字列判定工程と、

前記関連文字列判定工程によって前記検索条件に合致すると判定された場合にのみ、前記第1の文書構造データの次に配列されている第2の文書構造データが、前記検索条件に合致するか否かを判定する第2の文書構造データ判定工程と、

前記第2の文書構造データ判定工程によって前記検索条件に合致すると判定された場合にのみ、前記検索条件に合致した第2の文書構造データに関連付けられている前記文字列データ内の文字列を抽出する第2の関連文字列抽出工程と、

を含んだことを特徴とするXML文書検索方法。

【0080】

(付記4) 前記第1および第2の文書構造データは、前記ツリー構造を構成するノードの種類をあらわすノード型と、前記文字列データに対し関連付けするリンクデータとを含むことを特徴とする付記3に記載のXML文書検索方法。

【0081】

(付記5) 前記第1の文書構造データ判定工程および前記第2の文書構造データ判定工程は、前記ノード型および前記リンクデータが検索条件に合致するか否かを判定することを特徴とする付記4に記載のXML文書検索方法。

【0082】

(付記6) 前記第1および第2の関連文字列抽出工程は、前記リンクデータに基づいて、前記文字列データを抽出することを特徴とする付記4に記載のXML文書検索方法。

【0083】

(付記7) 複数の要素によって所定のツリー構造を構成するXML文書を入力させるXML文書入力工程と、

前記XML文書入力工程によって入力されたXML文書を構成する複数の要素を、要素の出現順に前記ツリー構造にしたがって一列に配列させる要素配列工程と、

前記要素配列工程によって一列に配列された各要素内における開始タグと終了タグとの間に記述されているタグ間文字列と、前記開始タグおよび前記終了タグ内に記述されているタグ内文字列と、からなる文字列データを抽出させる文字列

データ抽出工程と、

前記要素配列工程によって一列に配列された要素ごとに、前記ツリー構造上の種類をあらわすノード型を特定させるノード型特定工程と、

前記文字列データ抽出工程によって抽出された文字列データと、前記ノード型特定工程によって特定されたノード型とを前記要素ごとに関連付けるリンクデータを生成させるリンクデータ生成工程と、

をコンピュータに実行させることを特徴とするXML文書解析プログラム。

【0084】

(付記8) 所定のツリー構造を有するXML文書に記述されている文字列データの中から、検索対象となる文字列を検索させるXML文書検索プログラムであって、

前記検索対象となる文字列を検索するための検索条件を入力させる検索条件入力工程と、

前記XML文書を構成する複数の要素の出現順に前記ツリー構造にしたがって配列された第1の文書構造データが、前記検索条件に合致するか否かを判定させる第1の文書構造データ判定工程と、

前記第1の文書構造データ判定工程によって前記検索条件に合致すると判定された場合にのみ、前記検索条件に合致した第1の文書構造データに関連付けられている前記文字列データ内の文字列を抽出させる第1の関連文字列抽出工程と、

前記第1の関連文字列抽出工程によって抽出された文字列が、前記検索条件に合致するか否かを判定させる関連文字列判定工程と、

前記関連文字列判定工程によって前記検索条件に合致すると判定された場合にのみ、前記第1の文書構造データの次に配列されている第2の文書構造データが、前記検索条件に合致するか否かを判定させる第2の文書構造データ判定工程と、

前記第2の文書構造データ判定工程によって前記検索条件に合致すると判定された場合にのみ、前記検索条件に合致した第2の文書構造データに関連付けられている前記文字列データ内の文字列を抽出させる第2の関連文字列抽出工程と、

をコンピュータに実行させることを特徴とするXML文書検索プログラム。

【0085】

(付記9) 複数の要素によって所定のツリー構造を構成するXML文書を入力するXML文書入力手段と、

前記XML文書入力工程によって入力されたXML文書を構成する複数の要素を、要素の出現順に前記ツリー構造にしたがって一列に配列する要素配列手段と、

前記要素配列手段によって一列に配列された各要素内における開始タグと終了タグとの間に記述されているタグ間文字列と、前記開始タグおよび前記終了タグ内に記述されているタグ内文字列と、からなる文字列データを抽出する文字列データ抽出手段と、

前記要素配列手段によって一列に配列された要素ごとに、前記ツリー構造上の種類をあらわすノード型を特定するノード型特定手段と、

前記文字列データ抽出手段によって抽出された文字列データと、前記ノード型特定工程によって特定されたノード型とを前記要素ごとに関連付けるリンクデータを生成するリンクデータ生成手段と、

を含んだことを特徴とするXML文書解析装置。

【0086】

(付記10) 所定のツリー構造を有するXML文書に記述されている文字列データの中から、検索対象となる文字列を検索するXML文書検索装置であって、

前記検索対象となる文字列を検索するための検索条件を入力する検索条件入力手段と、

前記XML文書を構成する複数の要素の出現順に前記ツリー構造にしたがって配列された複数の文書構造データが、前記検索条件にそれぞれ合致するか否かを判定する文書構造データ判定手段と、

前記文書構造データ判定手段によって前記検索条件に合致すると判定された場合にのみ、前記検索条件に合致した文書構造データに関連付けられている前記文字列データ内の文字列を抽出する関連文字列抽出手段と、

前記関連文字列抽出手段によって抽出された文字列が、前記検索条件に合致するか否かを判定する関連文字列判定手段と、を備え、

前記関連文字列抽出手段は、前記関連文字列判定手段によって抽出された文字列が、前記検索条件に合致すると判定され、かつ、前記文書構造データ判定手段によって、前記検索条件に合致すると判定された文書構造データの次に配列されている文書構造データが前記検索条件に合致すると判定された場合、前記検索条件に合致すると判定された文書構造データ以外の文書構造データに対して関連付けられている前記文字列データ内の文字列を、前記検索対象となる文字列として抽出することを特徴とするXML文書検索装置。

【0087】

【発明の効果】

以上説明したように、この発明によれば、文字列データの参照頻度を少なくすることにより、XML文書の検索の高速化を図ることができるという効果を奏する。

【図面の簡単な説明】

【図1】

この発明の実施の形態にかかるXML文書解析／検索装置のハードウェア構成を示すブロック図である。

【図2】

この発明の実施の形態にかかるXML文書解析／検索装置の機能的構成を示すブロック図である。

【図3】

この発明の実施の形態にかかるXML文書のツリー構造を示す説明図である。

【図4】

この発明の実施の形態にかかるXML文書解析処理手順を示すフローチャートである。

【図5】

この発明の実施の形態にかかるXML文書解析処理におけるXML文書の構造配列を示す模式図である。

【図6】

この発明の実施の形態にかかるXML文書解析処理におけるXML文書からの

文字列データ生成処理を示す模式図である。

【図 7】

この発明の実施の形態にかかる XML 文書の一例を示す説明図である。

【図 8】

この発明の実施の形態にかかる文書構造データの一例を示す説明図である。

【図 9】

この発明の実施の形態にかかる文字列データ（名前空間データ）を示す説明図である。

【図 10】

この発明の実施の形態にかかる文字列データ（要素名データ）を示す説明図である。

【図 11】

この発明の実施の形態にかかる文字列データ（要素名／属性データ（整数データ））を示す説明図である。

【図 12】

この発明の実施の形態にかかる文字列データ（要素名／属性データ（テキストデータ））を示す説明図である。

【図 13】

この発明の実施の形態にかかる XML 文書検索処理手順を示すフローチャートである。

【図 14】

この発明の実施の形態にかかる XML 文書検索処理における検索条件の一例を示す説明図である。

【図 15】

この発明の実施の形態にかかる XML 文書検索処理における検索データの一例を示す説明図である。

【図 16】

この発明の実施の形態にかかる XML 文書検索実行処理手順を示すフローチャートである。

【図 1 7】

この発明の実施の形態にかかるXML文書検索実行処理における状態遷移を示す説明図である。

【符号の説明】

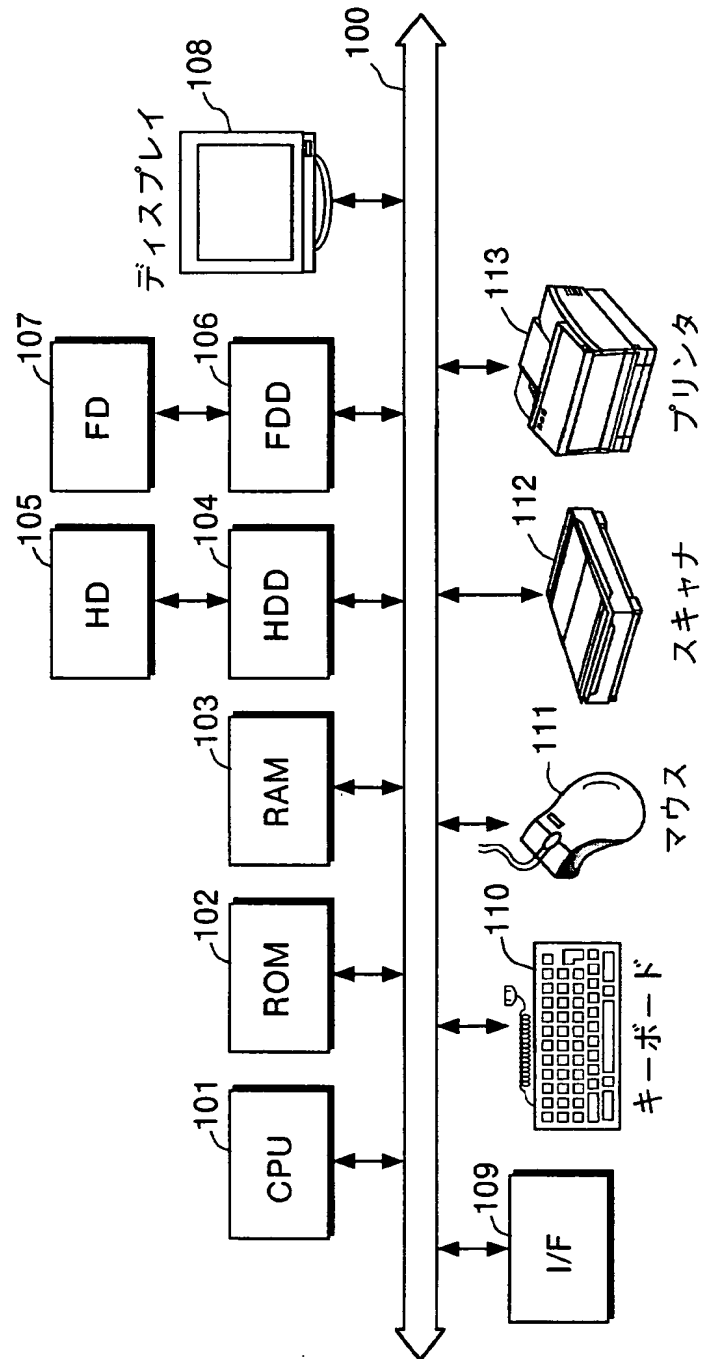
- 1 ルート要素
- 1 A ルート開始要素
- 1 B ルート閉じ要素
- 2 ～ 8 要素
- 4 A 開始要素
- 4 B 閉じ要素
- 2 0 0 XML文書解析／検索装置
- 2 0 1 XML文書入力部
- 2 0 2 XML文書解析部
- 2 0 3 要素配列部
- 2 0 4 文字列データ抽出部
- 2 0 5 ノード型特定部
- 2 0 6 リンクデータ生成部
- 2 0 8 XML文書格納部
- 2 1 1 検索条件入力部
- 2 1 3 XML文書検索部
- 2 1 4 ノード型走査部
- 2 1 5 ノード型判定部
- 2 1 6 関連文字列抽出部
- 2 1 7 関連文字列判定部
- 2 1 8 検索結果出力部

【書類名】

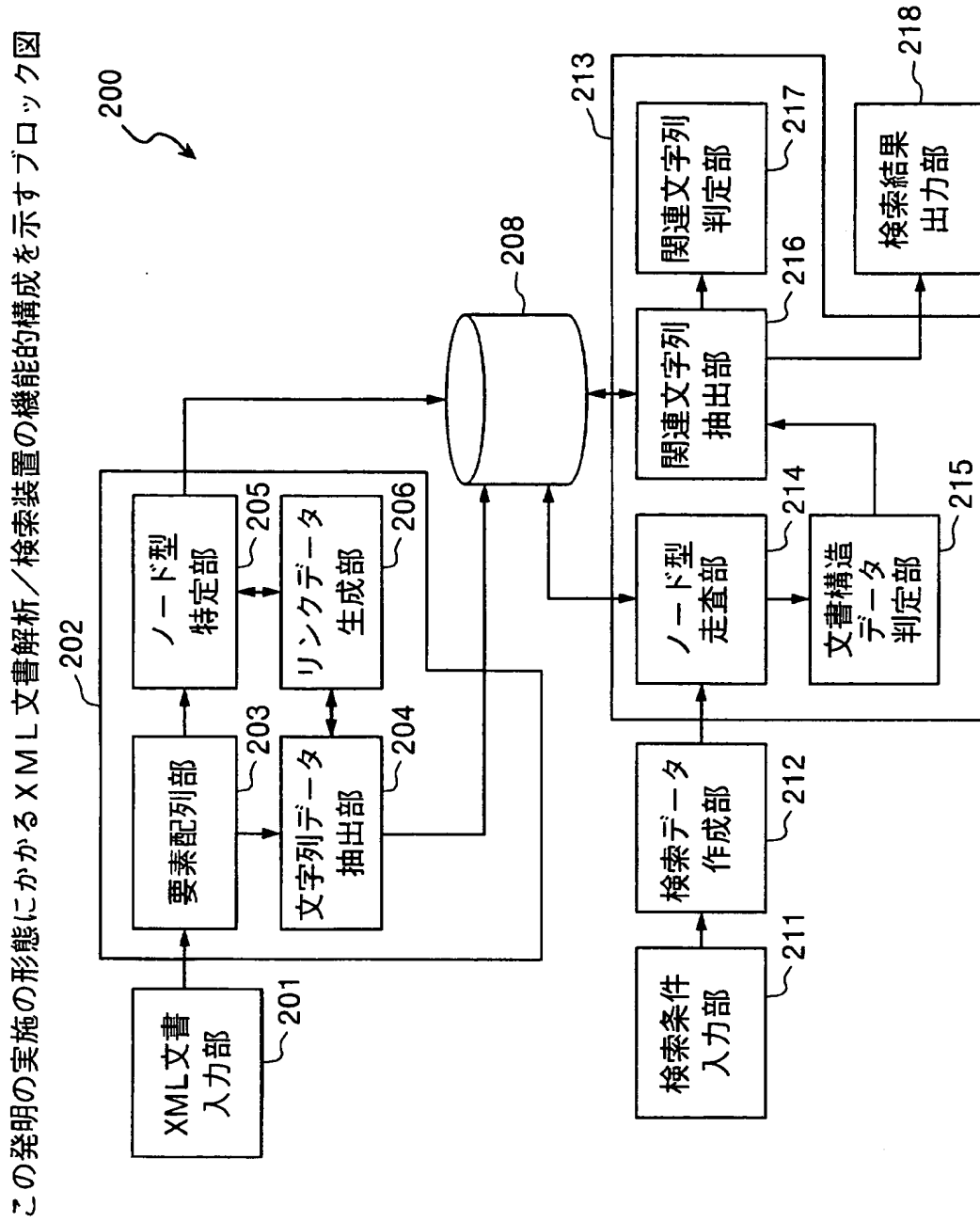
【図 1】

図面

この発明の実施の形態にかかるXML文書解析／検索装置のハードウェア構成を示すブロック図

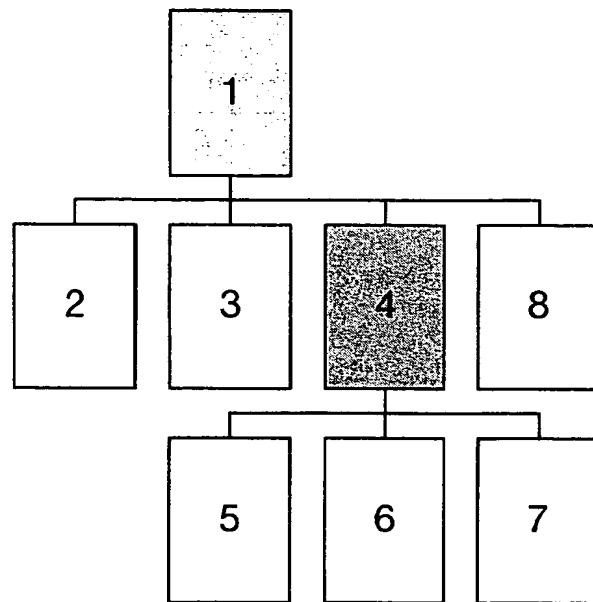


【図 2】



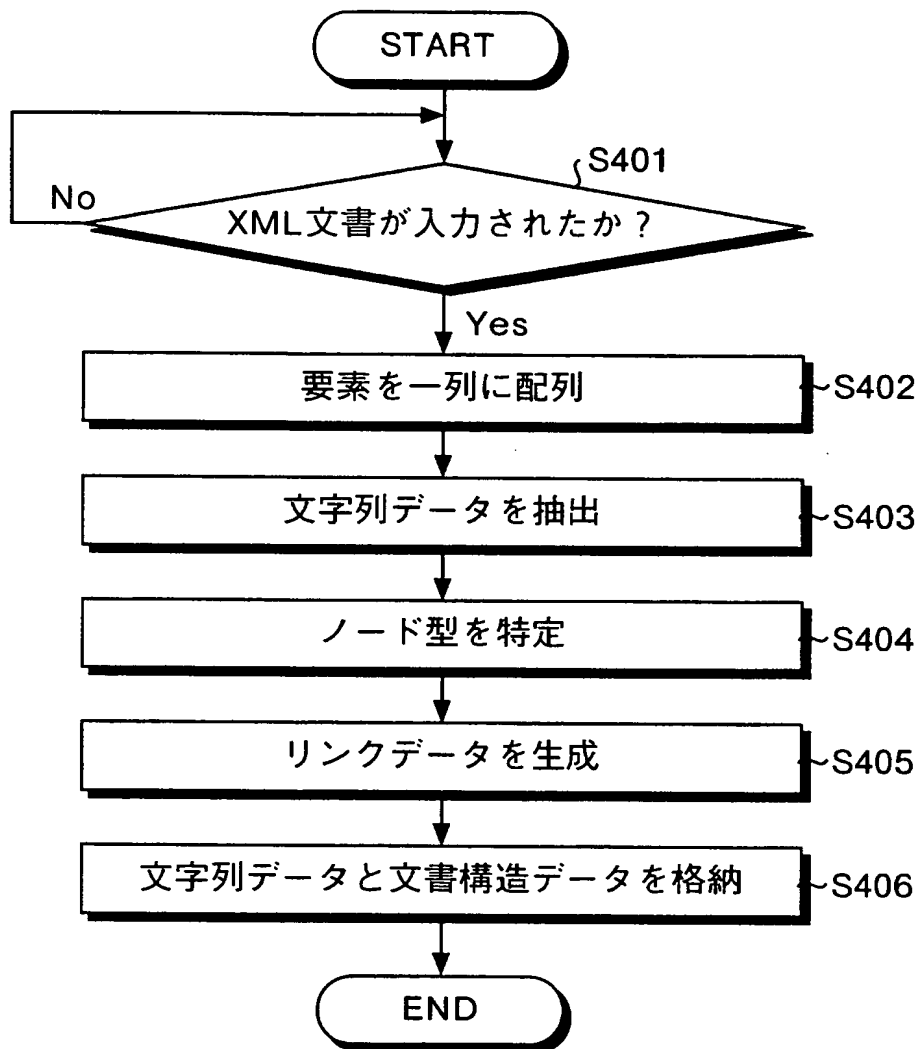
【図 3】

この発明の本実施の形態にかかるXML文書の
ツリー構造を示す説明図



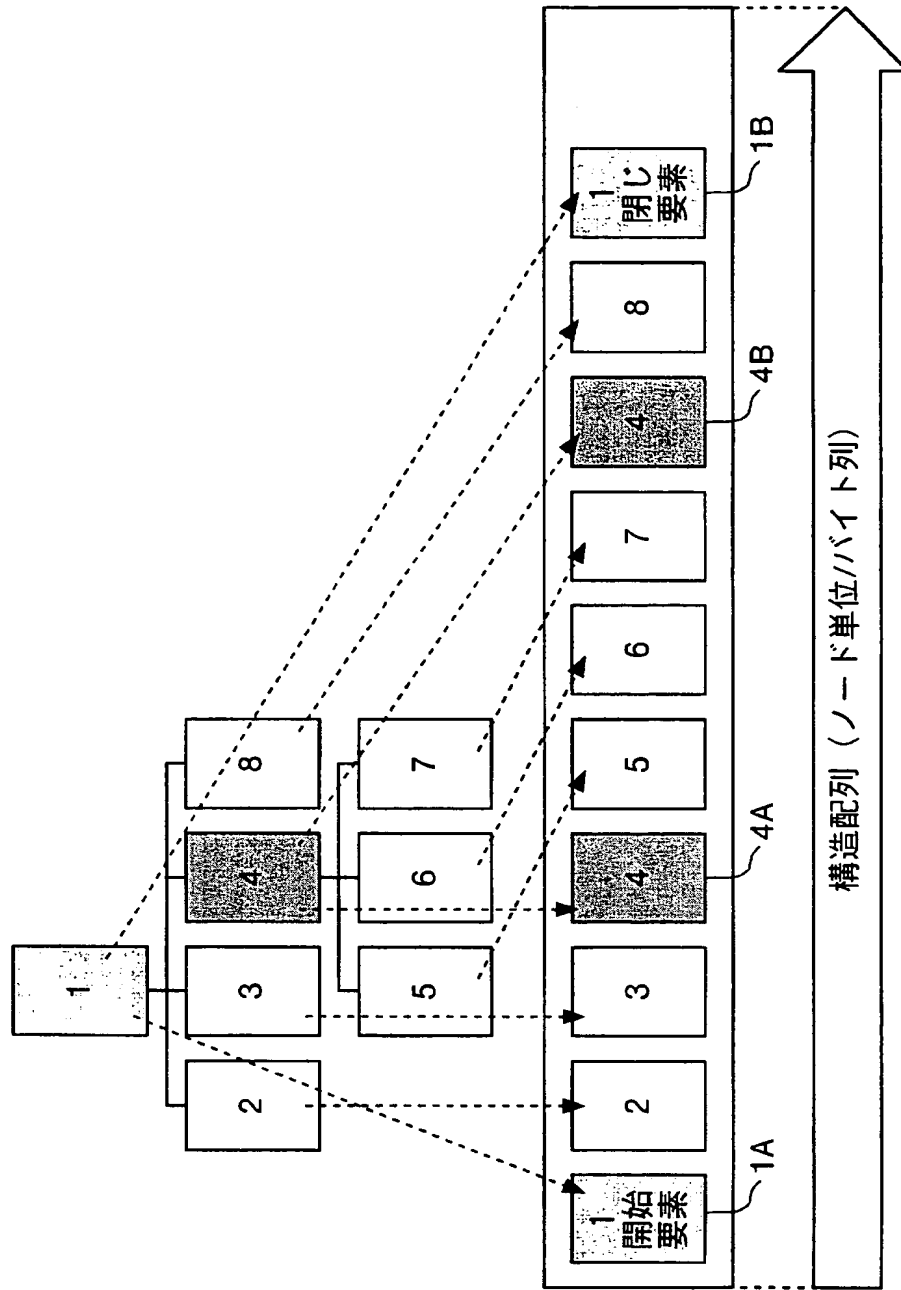
【図 4】

この発明の実施の形態にかかるXML文書解析処理手順を示す
フローチャート



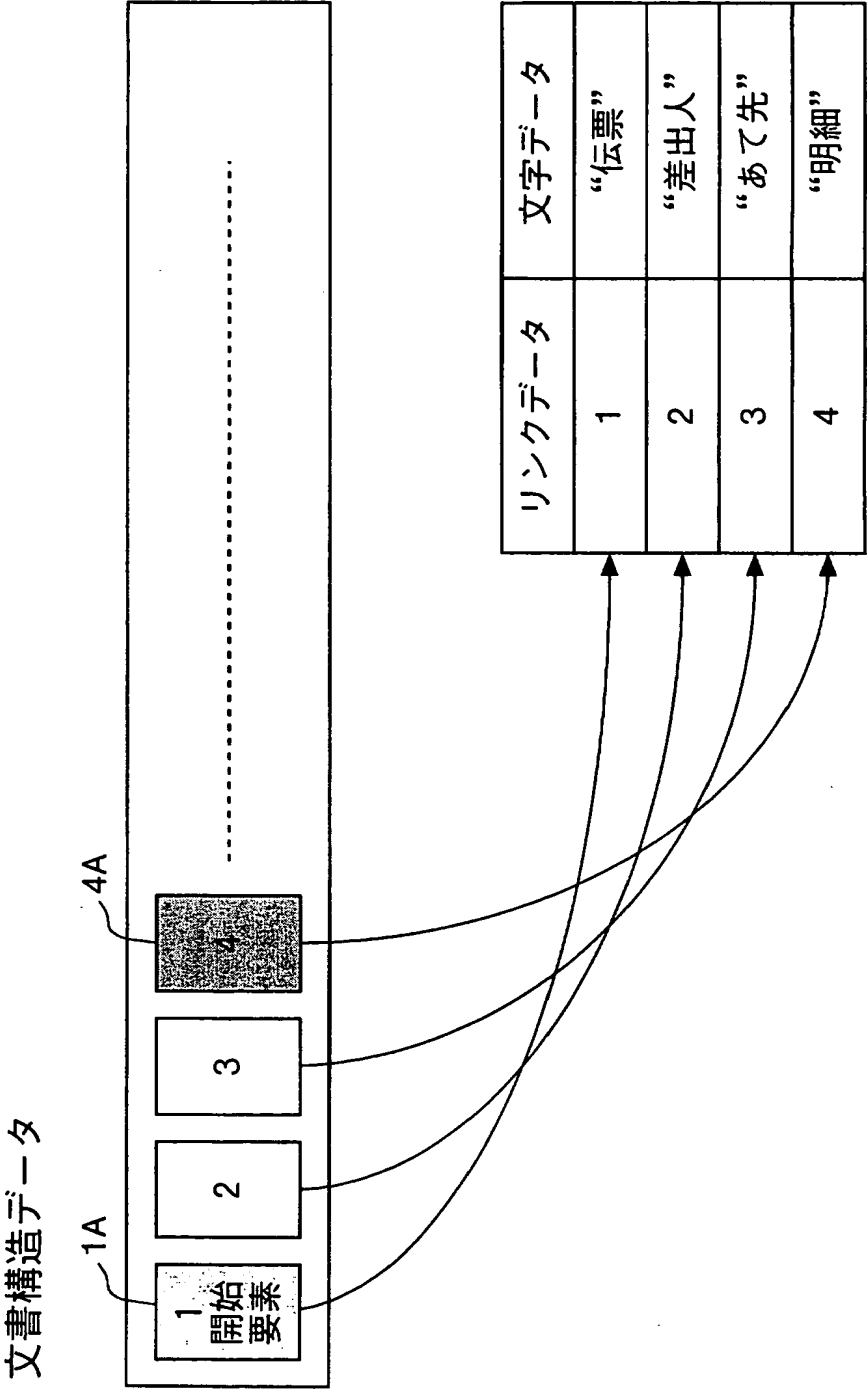
【図 5】

この発明の実施の形態にかかるXML文書解析処理におけるXML文書の構造配列を示す模式図



【図 6】

この発明の実施の形態にかかるXML文書解析処理におけるXML文書からの
文字列データ生成処理を示す模式図



【図 7】

この発明の実施の形態にかかるXML文書の一例を示す説明図

```
<製品リスト xmlns="urn:fj">  
  <製品 型番='0001'>  
    <名前>FMV</名前>  
    <価格>300000</価格>  
  </製品>  
  <製品 型番='0002'>  
    <名前>LOOX</名前>  
    <価格>200000</価格>  
  </製品>  
</製品リスト>
```

【図 8】

この発明の実施の形態にかかる文書構造データの一例を示す説明図

ノード型	リンクデータ			
	名前空間 ID	要素名 ID	データ型 ID	要素名/属性 データID
ドキュメント開始	—	—	—	—
要素開始	0	1	—	—
名前空間マップ	0	0	テキスト	0
要素開始	0	2	—	—
属性	0	3	テキスト	1
1 子要素	0	4	テキスト	2
1 子要素	0	5	整数	0
要素終了	0	2	—	—
要素開始	0	2	—	—
属性	0	3	テキスト	3
1 子要素	0	4	テキスト	4
1 子要素	0	5	整数	1
要素終了	0	2	—	—
要素終了	0	1	—	—
ドキュメント終了	—	—	—	—

【図 9】

この発明の実施の形態にかかる文字列データ
(名前空間データ) を示す説明図

リンクデータ	名前空間データ
名前空間ID	
0	Urn:fj

【図 10】

この発明の実施の形態にかかる文字列データ
(要素名データ) を示す説明図

リンクデータ	要素名データ
要素名ID	
0	—
1	製品リスト
2	製品
3	型番
4	名前
5	価格

【図 11】

この発明の実施の形態にかかる文字列データ
 (要素名/属性データ (整数データ)) を示す説明図

リンクデータ	要素名/属性データ
要素名/属性データID	整数データ
0	300000
1	200000

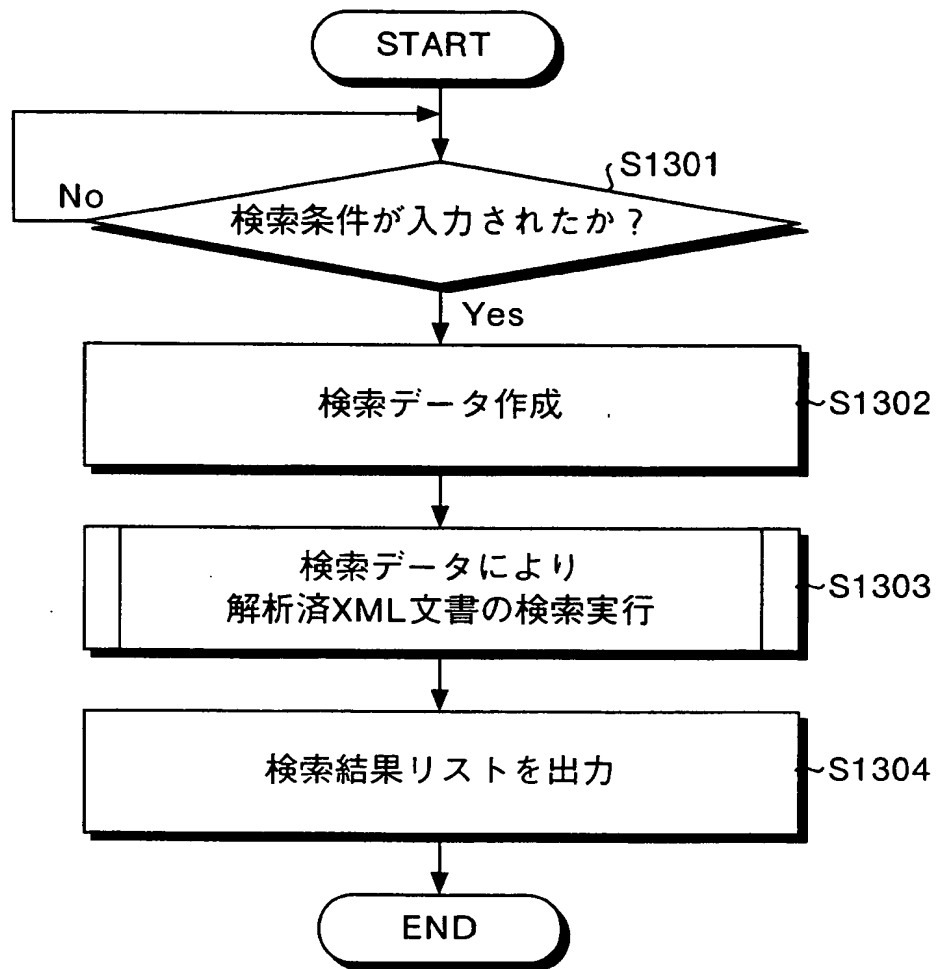
【図 12】

この発明の実施の形態にかかる文字列データ
 (要素名/属性データ (テキストデータ)) を示す説明図

リンクデータ	要素名/属性データ
要素名/属性データID	テキストデータ
0	Urn:fj
1	0001
2	FMV
3	0002
4	LOOX

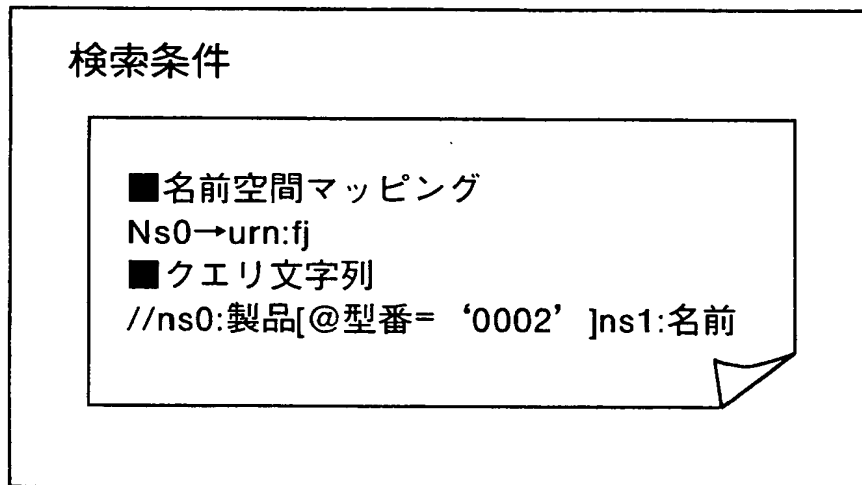
【図 13】

この発明の実施の形態にかかる XML 文書検索処理手順を
示すフローチャート



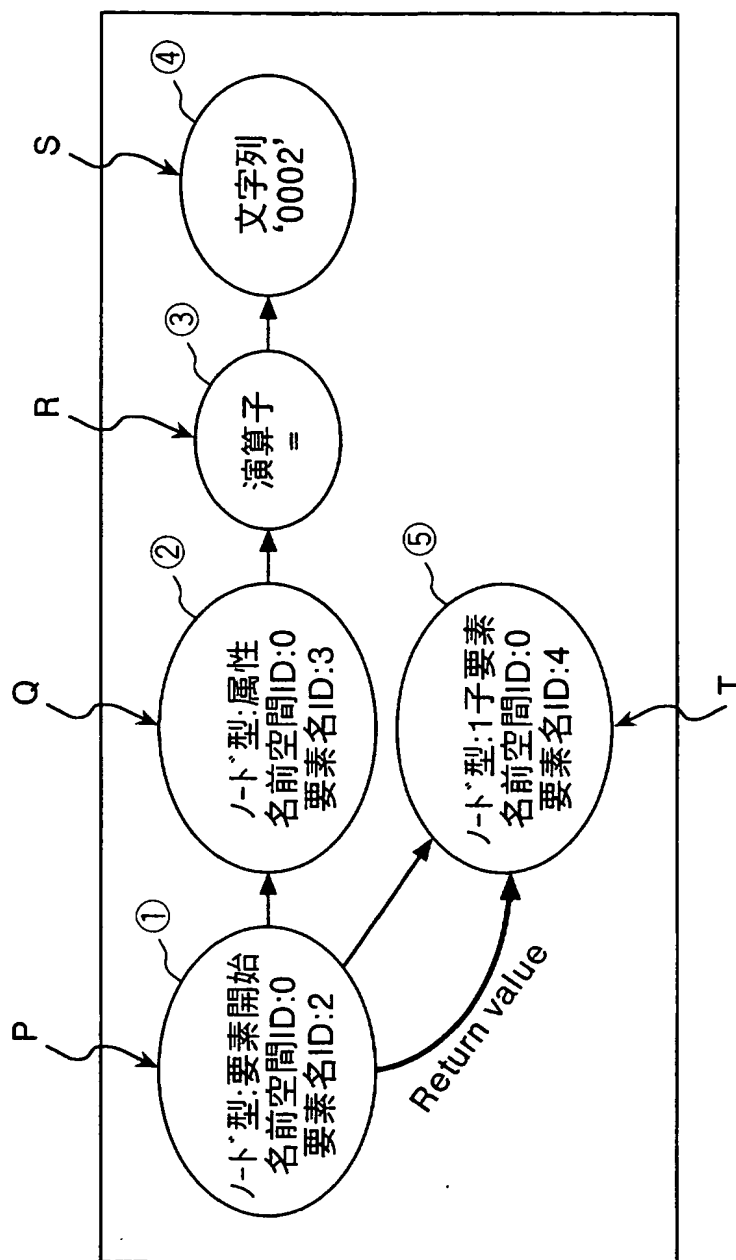
【図 14】

この発明の実施の形態にかかるXML文書検索処理における
検索条件の一例を示す説明図



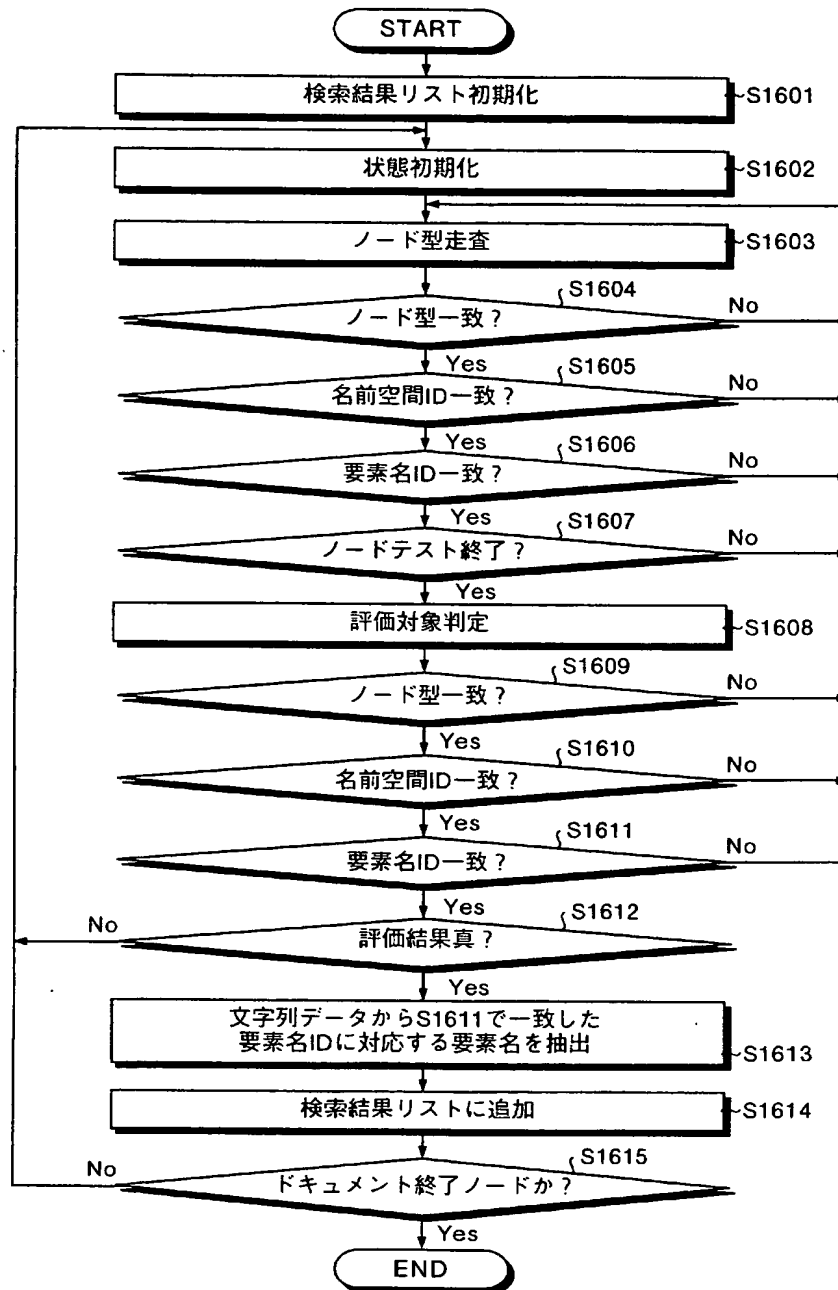
【図 15】

この発明の実施の形態にかかるXML文書検索処理における
検索データの一例を示す説明図



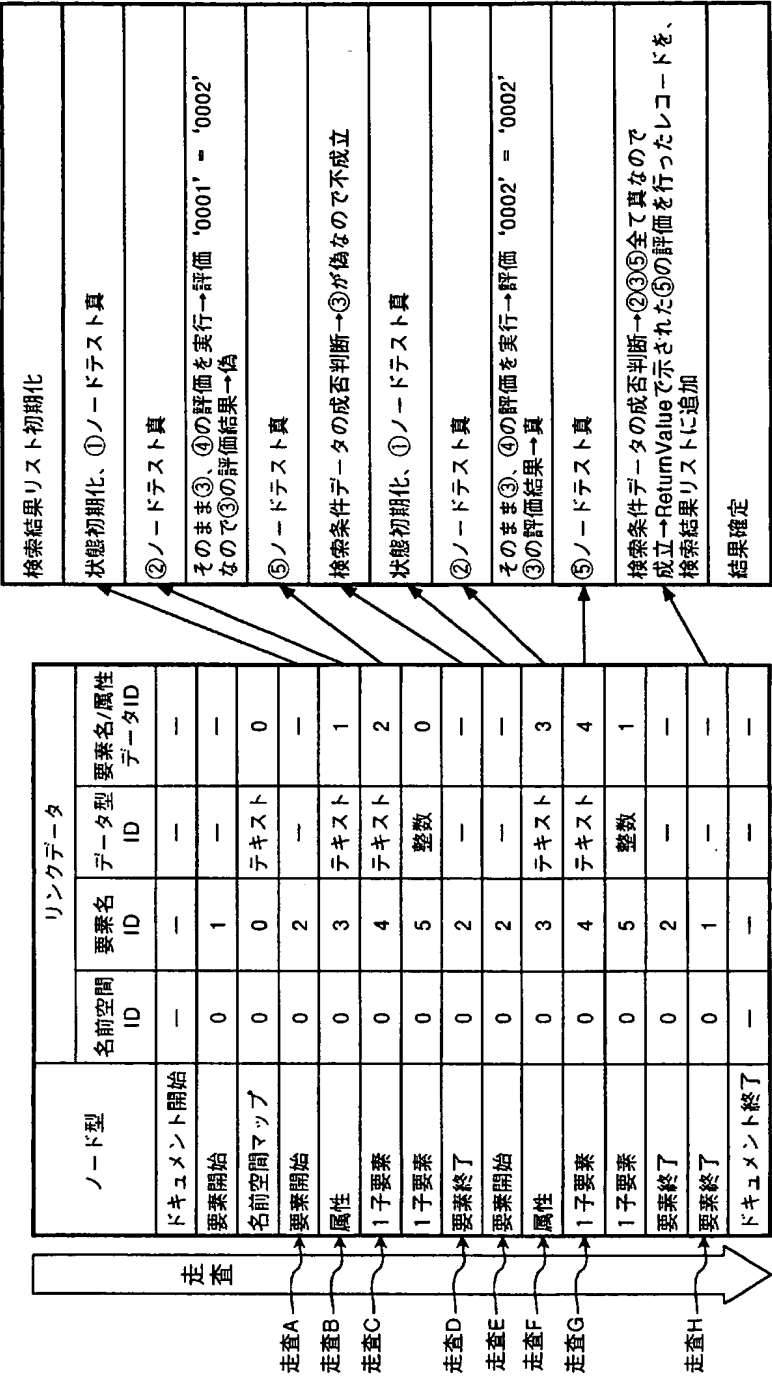
【図 16】

この発明の実施の形態にかかるXML文書検索実行処理手順を示すフローチャート



【図 17】

この発明の実施の形態にかかるXML文書検索実行処理における状態遷移を示す説明図



【書類名】 要約書

【要約】

【課題】 XML 文書の検索の高速化を図ること。

【解決手段】 XML 文書検索装置 200 は、検索条件を入力する検索条件入力部 211 と、XML 文書を構成する複数の要素の出現順にツリー構造にしたがって配列された複数の文書構造データが、検索条件にそれぞれ合致するか否かを判定する文書構造データ判定部 215 と、文書構造データ判定部 215 によって検索条件に合致すると判定された場合にのみ、検索条件に合致した文書構造データに関連付けられている文字列データ内の文字列を抽出する関連文字列抽出部 216 と、関連文字列抽出部 216 によって抽出された文字列が、検索条件に合致するか否かを判定する関連文字列判定部 217 と、を備える。

【選択図】 図 2

特願 2 0 0 3 - 0 8 9 1 2 6

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 2 2 3]

1 . 変更年月日
[変更理由]

1 9 9 6 年 3 月 2 6 日

住所変更

住 所
氏 名

神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号
富士通株式会社